# Peringkasan Multidokumen Otomatis dengan Menggunakan *Log-Likelihood Ratio* (LLR) dan *Maximal Marginal Relevance* (MMR) untuk Artikel dengan Topik Penyakit Menular Bahasa Indonesia

**SKRIPSI**

**Diajukan untuk Memenuhi Salah Satu Syarat Mencapai Gelar Strata Satu Program Studi Informatika**


Disusun oleh :

**IKHWAN NIZWAR AKHMAD**

**M0511025**



**PROGRAM STUDI INFORMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS SEBELAS MARET**

**SURAKARTA**


**2017**

**SKRIPSI**

**Peringkasan Multidokumen Otomatis dengan *Menggunakan Log-Likelihood Ratio* (LLR) dan *Maximal Marginal Relevance* (MMR) untuk Artikel dengan Topik Penyakit Menular Bahasa Indonesia**

**Disusun Oleh :**

**IKHWAN NIZWAR AKHMAD**

**M0511025**

**Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji pada tanggal,**

| **Pembimbing I** | **Pembimbing II** |
|---|---|
| **Drs. Bambang Harjito M.App.Sc., Ph.D** | **Dr. Eng. Anto Satriyo Nugroho** |
| **NIP. 19621130 199103 1 002** | **NIP. 19701021 198911 1 001** |

**SKRIPSI**

**Peringkasan Multidokumen Otomatis dengan Menggunakan Log-Likelihood Ratio (LLR) dan Maximal Marginal Relevance (MMR) untuk Artikel dengan Topik Penyakit Menular Bahasa Indonesia**

**Disusun Oleh:**

**IKHWAN NIZWAR AKHMAD**

**M0511025**

**Telah dipertahankan dihadapan Dewan Penguji**

**pada tanggal, 15 Desember 2017**

**Susunan Dewan Penguji**

1. **Drs. Bambang Harjito M.App.Sc., Ph.D.** **(Ketua)** **( )**
   **NIP. 196211301991031002**
2. **Dr. Anto Satriyo Nugroho, M. Eng.** **(Sekretaris)** **( )**
   **NIP. 197010211989111001**
3. **Sari Widya Sihwi, S.Kom., M.T.I.** **(Anggota)** **( )**
   **NIP. 198304122009122003**
4. **Denis Eka Cahyani, S.Kom, M.Kom.** **(Anggota)** **( )**
   **NIP. 1991031020161001**

**Disahkan Oleh**

**Kepala Program Studi Informatika**

**Drs. Bambang Harjito M.App.Sc., Ph.D.**

**NIP. 196211301991031002**

## HALAMAN MOTTO

“Bila engkau ingin satu, maka jangan ambil dua. Karena satu menggenapkan, tapi dua melenyapkan.” - Dee

# HALAMAN PERSEMBAHAN

Untuk Abi, Umik, Hani, Iin, dan Sekar.

# KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "Peringkasan Multidokumen Otomatis dengan Menggunakan *Log-Likelihood Ratio* (LLR) dan *Maximal Marginal Relevance* (MMR) untuk Artikel dengan Topik Penyakit Menular Bahasa Indonesia". Skripsi ini disusun sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer pada Program Studi Informatika Universitas Sebelas Maret.

Dalam proses penelitian dan penyusunan skripsi ini, penulis telah mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Untuk itu, penulis menyampaikan terima kasih kepada:

1. Bapak, Ibu, serta Adik yang telah dan terus-menerus memberikan dukungan material dan non-material selama proses penyelesaian tugas akhir.

2. Bapak Dr. Anto Satriyo Nugroho, M.Eng selaku Pembimbing Tugas Akhir yang telah memberikan banyak solusi atas permasalahan penulis selama proses penyelesaian tugas akhir, serta membuka wawasan penulis mengenai ranah penelitian di Indonesia.

3. Bapak Drs. Bambang Harjito, M.APP.Sc, Ph.D. selaku Kepala Program Studi Informatika Universitas Sebelas Maret, Pembimbing Akademik, dan Pembimbing Tugas Akhir, yang telah memberikan banyak motivasi, saran, dan bimbingan selama proses studi dan penyelesaian tugas akhir.

4. Bapak Prof. Ir. Ari Handono Ramelan, M.Sc.(Hons), Ph.D. selaku Dekan Fakultas MIPA Universitas Sebelas Maret.

5. Seluruh Dosen Program Studi Informatika yang telah membagikan begitu banyak dasar ilmu pengetahuan dan pengalamannya kepada penulis.

6. Teman-teman Informatika dan Ngarsapura Creative Media, yang telah menjadi bagian dari perjalanan studi, pembentukan karakter diri, juga susah senang sebagai mahasiswa dan rekan kerja.

Kesempurnaan memang hanya milik Allah SWT. Namun Penulis juga menyadari bahwa penelitian dan skripsi ini masih jauh dari penelitian dan laporan yang ideal. Meskipun demikian, penulis sangat berharap skripsi ini bisa bermanfaat bagi pembaca.

Surakarta, November 2017

Penulis

**Peringkasan Multidokumen Otomatis dengan Menggunakan *Log-Likelihood Ratio* (LLR) dan *Maximal Marginal Relevance* (MMR) untuk Artikel dengan Topik Penyakit Menular Bahasa Indonesia**

**Ikhwan Nizwar Akhmad**

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

## ABSTRAK

Peningkatan jumlah informasi yang tersedia di internet disamping memberikan manfaat, juga memunculkan masalah tersendiri. Mesin pencarian modern sudah cukup baik untuk mendapatkan informasi tertentu. Namun jumlah informasi yang sangat banyak terkadang menyebabkan pencari informasi kesulitan mendapatkan intisari dari informasi yang dicari. Kondisi yang disebabkan oleh terlalu banyaknya informasi yang tersedia dikenal sebagai *information overload*.

Peringkasan multidokumen otomatis merupakan salah satu solusi untuk masalah ini. Meskipun metode peringkasan multidokumen otomatis sudah dikembangkan sejak 20 tahun lalu, penerapannya dalam Bahasa indonesia masih terbatas. Artikel dengan topik penyakit menular merupakan salah satu studi kasus yang menarik untuk peringkasan multidokumen Bahasa Indonesia. Informasi mengenai penyakit menular dibutuhkan oleh masyarakat sehingga terdapat banyak informasi mengenai topik ini di internet. Kondisi ini menyebabkan kemungkinan *information overload* untuk pencarian dalam topik ini.

Dalam penelitian ini, akan diterapkan metode peringkasan multidokumen otomatis dengan menggunakan *Log-Likelihood Ratio* (LLR) untuk mendapatkan *topic signature*, dan *Maximal Marginal Relevance* pada artikel dengan topik penyakit menular untuk mendapatkan ringkasan dengan sedikit perulangan informasi.

Penelitian ini menghasilkan ringkasan dengan nilai akurasi sebesar 0,4 (dengan menggunakan ROUGE-S9). Selain itu, dalam penelitian ini didapatkan bahwa *topic signature* (beserta akurasinya) memegang peran penting dalam proses peringkasan dokumen otomatis.

**Kata Kunci:** peringkasan multidokumen otomatis, *topic signature generation*

***Automatic Multidocument Summarization utilizing Log-Likelihood Ratio (LLR) and Maximal Marginal Relevance (MMR) for Infetious Diseases Articles in Indonesian***

**Ikhwan Nizwar Akhmad**

*Department of Informatics, Mathematics and Natural Science Faculty,*

*Sebelas Maret University*

## ABSTRACT

*Increasing number of information that available on the Internet, along with its benefit, also comes with its own problems. Modern search engines are smart enough to bring the most corresponding information, but the immense number of information provided sometimes bring more confusion than clarity. This condition is known as information overload.*

*Automatic multidocument summarization is a way to overcome this particular problem. But despite being heavily researched more than 20 years ago, its implementations for Bahasa Indonesia are limited. Articles about infectious disease is one of the ideal case study for multidocument summarization for Bahasa Indonesia. Information about infectious disease are essential for general public therefore many information about it is available on the Internet. This condition could trigger information overload when someone do an internet search in this topic.*

*In this research, we try to implement multidocument summarization technique for articles with infectious disease topic in Bahasa Indonesia utilizing Log Likelihood Ratio (LLR) to obtain topic signatures and Maximal Marginal Relevance (MMR) to generate relevant summary with minimal information redundancy.*

*Our summarization method generated a summary with 0.4 fmeasure using ROUGE-S9 evalution. Also, we found that topic signature (with its accuracy) takes an important role on generating good summaries.*

***Keywords:*** *multidocument summarization, topic signatures generation*

## DAFTAR ISI

## DAFTAR GAMBAR

## DAFTAR TABEL